AD-A050 210 ILLINOIS UNIV AT URBANA-CHAMPAIGN SAVOY AVIATION RES--ETC F/G 5/8
THE EFFECTS OF PARTICIPATORY MODE AND TASK WORKLOAD ON THE DETE--ETC(U)
JUN 77 C D WICKENS, C KESSEL F44620-76-C-0009
UNCLASSIFIED ARL-77-8/AFOSR-77-7 AFOSR-TR-78-0137 NL

| OF |

AD
A050210

END
DATE
FILMED
3 -78
DDC

AFOSR-TR- 78-0137

AD A050210

AVIATION RESEARCH LABORATORY

INSTITUTE OF AVIATION
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

TECHNICAL REPORT

THE EFFECTS OF PARTICIPATORY MODE AND TASK WORKLOAD ON THE DETECTION OF DYNAMIC SYSTEM FAILURES

CHRISTOPHER D. WICKENS, COLIN KESSEL

ARL-77-8/AFOSR-77-7
JUNE 1977

F44620-76-C-0009

Approved for public release;
distribution unlimited.

PREPARED FOR
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFSC)
NOTICE OF TRANSMITTAL TO DDC
This technical report has been reviewed and is
approved for public release IAW AFR 190-12 (7b).
Distribution is unlimited.
A. D. BLOSE
Technical Information Officer

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

19 REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: 18 AFOSR-TR-78-0137

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER: 9 Interim technical rept.,

4. TITLE (and Subtitle): 6 THE EFFECTS OF PARTICIPATORY MODE AND TASK WORKLOAD ON THE DETECTION OF DYNAMIC SYSTEM FAILURES.

5. TYPE OF REPORT & PERIOD COVERED: ~~Scientific Interim~~

6. PERFORMING ORG. REPORT NUMBER: 14 ARL-77-8/AFOSR-77-7

7. AUTHOR(s): 10 Christopher D./Wickens and Colin/Kessel

8. CONTRACT OR GRANT NUMBER(s): 15 F44620-76-C-0009

9. PERFORMING ORGANIZATION NAME AND ADDRESS: Aviation Research ~~Laboratory~~ ~~Institute of Aviation~~ University of Illinois, Savoy, Illinois 61874

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 61102F 16 2313/A4 17 A4

11. CONTROLLING OFFICE NAME AND ADDRESS: Air Force Office of Scientific Research (NL) Building 410 - Bolling AFB Washington, DC 20332

12. REPORT DATE: 11 Jun[redacted]977

13. NUMBER OF PAGES: 38 12 46P.

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

15. SECURITY CLASS. (of this report): Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Dynamic Systems, Internal model, Failure detection, Workload

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The ability of operators to detect step changes in the order of control dynamics is investigated as a joint function of a) participatory mode: whether subjects are actively controlling those dynamics or are monitoring an autopilot controlling them, and b) concurrent task workload. A theoretical analysis of detection in the two modes identifies factors that will favor detection in either mode. Five subjects either tracked or monitored the system dynamics on a 2-dimensional pursuit display under single task conditions and concurrently over

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

406171 JOB

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

20.
with a "subcritical" tracking task at two difficulty levels. Latency and accuracy of detection were assessed and related through a speed-accuracy tradeoff representation. Detection performance was faster, and only slightly less accurate in the manual as opposed to the autopilot mode and performance in each mode was derogated by the concurrent tracking requirement, but not by increases in loading task difficulty. Further analysis, involving multiple regression techniques, ensemble averaging and examination of response latency distributions suggested that manual superiority was attributable to the additional proprioceptive information resulting from control adaption to the system change. The effects of the loading task on detection and upon primary task tracking were interpreted in terms of the concept of limited processing resources.

ABSTRACT

The ability of operators to detect step changes in the order of control dynamics is investigated as a joint function of a) participatory mode: whether subjects are actively controlling those dynamics or are monitoring an autopilot controlling them, and b) concurrent task workload. A theoretical analysis of detection in the two modes identifies factors that will favor detection in either mode. Five subjects either tracked or monitored the system dynamics on a 2-dimensional pursuit display under single task conditions and concurrently with a "subcritical" tracking task at two difficulty levels. Latency and accuracy of detection were assessed and related through a speed-accuracy tradeoff representation. Detection performance was faster, and only slightly less accurate in the manual as opposed to the autopilot mode and performance in each mode was derogated by the concurrent tracking requirement, but not by increases in loading task difficulty. Further analysis, involving multiple regression techniques, ensemble averaging and examination of response latency distributions suggested that manual superiority was attributable to the additional proprioceptive information resulting from control adaption to the system change. The effects of the loading task on detection and upon primary task tracking were interpreted in terms of the concept of limited processing resources.

AVIATION RESEARCH LABORATORY

Institute of Aviation

University of Illinois at Urbana-Champaign
Willard Airport
Savoy, Illinois
61874

Technical Report

ARL-77-8/AFOSR-77-7

June 1977

THE EFFECTS OF PARTICIPARTORY MODE AND TASK WORKLOAD ON THE DETECTION OF DYNAMIC SYSTEM FAILURES

Christopher D. Wickens

Colin Kessel

Prepared for

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE

F44620-76-C-0009

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Over the past decade, the aviation industry has witnessed a gradual change in the role of the pilot in the cockpit. Many traditional pilot functions have been replaced by on-board computers, and in some instances the pilot is no more than a supervisor [1] or monitor of automatically controlled functions. One task, however, that remains of critical importance to the operator of any aviation system, whether he is removed from the control loop or not, is that of monitoring all facets of aircraft performance for the occurrence of failures or malfunctions. The relatively low frequency of occurrence of such events does not diminish the importance of failure monitoring and detection, because the consequences of an undetected malfunction, or one that is detected after an unnecessary delay, can be disastrous, potentially resulting in the loss of the aircraft or of human life. It can be argued in fact that one criterion that should be used in considering whether a pilot should remain in the control loop under particular conditions is his relative sensitivity to system malfunctions in the two modes of participation.

Young [2] has argued strongly on the basis of his findings that the operator is more sensitive to system malfunctions as an active participant in the control loop, than as a passive monitor. In his experiment, subjects were required to detect various step changes in system order and gain. Conditions were compared in which the subject was an active controller and a passive monitor (who was observing the compensatory display produced by another active controller). Under these circumstances detection latencies were two to five times greater for the monitor than the controller. A second study which also compared detection ability in the two modes,however, resulted in contradictory findings.

Ephrath [3] investigated failure detection performance in a two-dimensional simulated landing task as a joint function of participatory mode and workload. The "failures," which in this case were deviations introduced into the flight path

rather than changes in system dynamics, could occur in either the pitch or yaw channel. Under different conditions subjects were either in control when a failure occurred or were monitoring a nonadaptive autopilot in control of that channel. The non-failed channel could also be either controlled or monitored. Ephrath's results indicated a clear superiority for detection on the monitored as opposed to the controlled dimension, both in terms of the smaller number of missed failures and of the shorter detection latency. This difference Ephrath attributed in large part to the increased level of workload involved in the controlled task.

Obviously, in many respects the studies of Young and Ephrath are not comparable. Young employed single-axis tracking with changes in system dynamics, while Ephrath employed dual-axis simulator control with "deviation" failures together with a secondary task designed to measure workload. In addition, the monitoring conditions were different in the two experiments, being influenced by adaptation in Young's study and not in Ephrath's. In this light, it is not surprising that the conclusions differed dramatically. Certainly, one of the most salient differences between these studies lies in the contrast between single- and dual-axis tracking and is inherent in the greater workload imposed in the latter condition.

While numerous other investigations of failure detection performance are present in the literature [2,4], the studies of Young and Ephrath are the only two that have explicitly contrasted detection between the two modes, so that a direct comparison is possible. The present study was conducted with the intent of clarifying the nature of the superiority relation between the two modes. A question of specific interest was whether the difference in results between the results of Ephrath's and Young's study could be attributable to differences in concurrent task workload between the paradigms, and for this reason secondary task workload was manipulated orthogonally to participatory mode.

Theoretical Analysis of Failure Detection

The detection of a failure or change in the characteristics of a dynamic system requires that the detector have available two basic elements: (1) an internal representation of the state of the normally operating system--the expected value of state variables and their expected variability [5,6] and (2) a channel, or set of channels, of information concerning the current state of the system. Failures are detected when the information concerning the current system state is assessed to be sufficiently deviant from the representation of normal operation to warrent a decision. The decision process involved may be assumed to involve the application of some statistical decision rule [7].

More specifically it is assumed that the detection process involves the integration over time of noisy evidence concerning any differences between current and normal operation until the decision criterion is exceeded within some time period. Provided that a failure has occured, the quality of this noisy information will increase with the number of information sources and with integration time, as long as memory for the standard of normal operation remains salient. Provided memory is salient then, failure decisions should become more prevalent at longer latencies. However it is also reasonable to assert that, if sufficient time has lapsed with the post-failure dynamics, the internal model of normal operation itself begins to reflect partially the new dynamics, and thus the "strength" of the difference signal that is integrated becomes attenuated. Thus while integrated information grows over time, the diagnostic value of that information will eventually decline, dictating that detection accuracy (or number of detections) will not be monotonically increasing with latency but rather will reflect this tradeoff, achieving a maximum at an intermediate latency.

Figure 1 presents a schematic comparison between the failure detection process described when the operator is in the control loop (top) and when he is removed from the loop, monitoring autopilot control (bottom). In both modes,

the available channels of information to the decision making system concerning current system state are represented. The following theoretical analysis, employing the conceptual framework described above and represented symbolically in Figure 1, will attempt to define the characteristics or attributes of each participatory mode that might be expected to enhance the sensitivity of failure detection in that mode.

Insert Figure 1 about here

Consistency of internal model of dynamics. An evolving conception in control theory is that the operator maintains an internal representation or "model in the head" of the dynamic system that is being controlled. It is assumed here that this model provides the basis for predicting expected system outputs in response to known inputs - an internalized estimate of the transfer function of the system being controlled. This conception is consistent with that employed by Curry & Gai [7] and Miller and Elkind [8] and others. With respect to failure detection, a critical characteristic of an internal model relates to its internal consistency or expected variability. For any given input to the system, the range or variability of expected outputs is a measure of this consistency.

It is proposed that when the operator is actively controlling, the stability of this internal model is considerably greater than when he is monitoring. This difference reflects the fact that, when controlling, the operator has a greater involvement with the system, and a direct knowledge of its input-output characteristics available by comparing his control inputs with the system response [9]. This information is only available when monitoring if the monitored display is pursuit, and even then, knowledge of control inputs is not as precise, since a system response due to regulatory error correction, cannot easily be discriminated

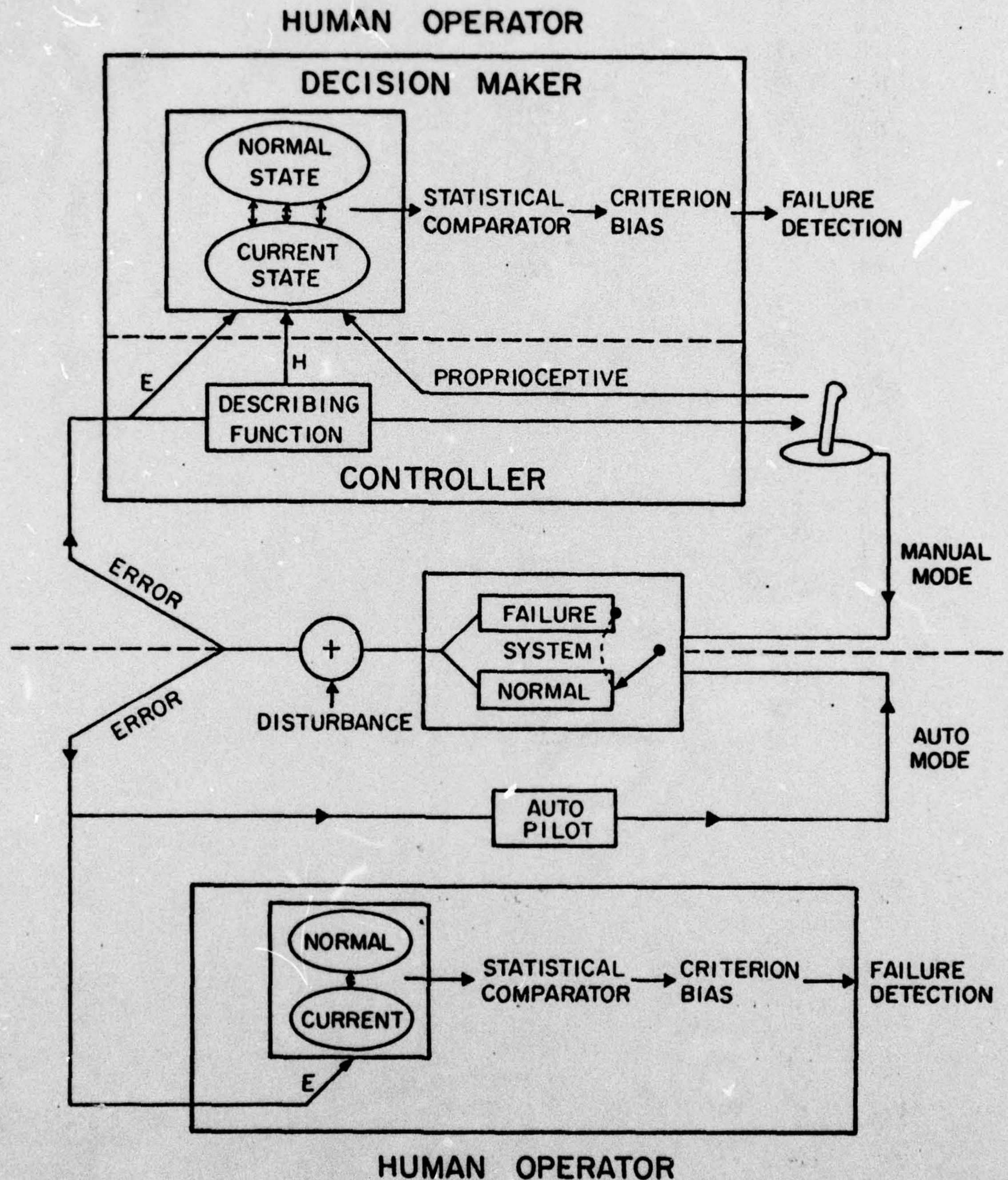

Figure 1: Schematic representation of failure detection process in control mode (top), monitor mode (bottom).

from one resulting from external disturbances. Thus in the control mode a smaller variance estimate of the normal state exists in the decision center and therefore detection should show greater sensitivity to departures from this state induced by changing dynamics, than should detection in the monitoring mode.

Information channels. A second attribute of the control mode that predicts superiority of failure detection is the greater number of channels of information concerning the current state. When monitoring, information is provided to the operator exclusively via the visual channel (system error and its derivatives in the compensatory display or input and outputs plus derivatives in the pursuit display). On the other hand, in the control situation the operator also has available a proprioceptive channel of information concerning his own input to the control stick, independent of disturbances acting upon the system.

Although control input cannot directly reflect the occurrence of failures (except as failures initiate mechanical feedback from the control itself), it will do so indirectly to the extent that any compensatory adaptation that the operator initiates to a system change will be reflected in a change in his response characteristics (mean control position, velocity or acceleration) and/or the characteristics of the operator's open-loop transfer function. When controlling then, these proprioceptive channels will be available to the detection system to supplement the visual channels that are available in both monitoring and controlling (Figure 1).

While the controlling mode thus seemingly provides a distinct advantage over the monitoring mode by virtue of its added proprioceptive channel, it should be noted that this advantage is not invariably present for reasons related to the non-independence of control input and error. More specifically, if adaptation to the failure is rapid and complete, as may occur for example in response to shifts in system gain [2], the obtained distribution of error following the change would show little or no alteration from that characterizing the normal

operating state, while a change would be manifest in the characteristics of the control response and, therefore, the transfer function.

Failure to initiate any adaptive control, on the other hand, would leave unchanged the proprioceptive input, while altering both the nature of the error distribution and again, the resulting transfer function. In short, whether or not an adaptive response is implemented, the transfer function will change. If adaptation occurs, the response will change as well. If it does not, then the error distribution will be altered.

However, even provided with only two sources of information (transfer function plus error _or_ control response) rather than three, a comparison of number of channels still favors the control mode over monitoring. Assuming that there is some degree of independence of information processing along the channels, the probability and/or speed of detecting change information along any one of two channels characterizing the control mode, should be greater than that of detecting change along the single visual channel available in the monitoring mode.

Differential sensitivity to visual vs proprioceptive information. Although a strict comparison of the number of channels of information available to a decision mechanism favors control over monitoring, an important caution should be noted. As described above, the operator is able to trade off the strength of the failure occurrence "signal" along the visual vs proprioceptive channel, to the extent that he engages in some degree of compensatory adaptation. As adaptation increases, proprioceptive "signal strength" increases at the expense of visual error "signal strength." Thus the prediction based upon the difference in number of information channels -- that control detection will be superior to monitoring detection -- is predicated upon the assumption that detection of change is equally efficient along all the channels (proprioceptive, transfer function, and visual). In other words this approach assumes that, _whichever_ channels are employed in the control mode, their joint signal will be more easily detected than the single visual signal in the monitoring mode.

Mitigating against this conclusion, however, is a body of literature in psychology suggesting that the sensitivity to proprioceptive information is reduced relative to visual information particularly when the two sources are available at the same time and are conveying conflicting information [e.g., 10, 11, 12]. Such a conflict, in fact, describes precisely the situation in which an operator has successfully adapted to a change in control dynamics. Under these circumstances, the visual error channel is providing information describing normal operation (since the appropriate gain, or lead-lag adjustment, has presumably been initiated to restore the original open-loop transfer characteristics), while the less sensitive kinesthetic channel conveys the information that a change has in fact been implemented. The predicted consequence of this conflict situation is that the operator will be less likely to detect the change than he would had no adaptation been achieved, the latter condition of course producing a visual signal equivalent to the monitoring mode. McDonnell [13], in fact, has noted anecdotally such instances in which successful adaptation has been coupled with the failure to detect dynamic system changes.

Workload differences. A second characteristic of the manual control mode that predicts a reduced sensitivity to the occurrence of failures relates to the greater workload imposed by tracking than by monitoring. Numerous examples may be cited from behavioral literature that demonstrate the attention demands of purely perceptual tasks such as monitoring to be less than those of tasks such as tracking in which a requirement for the selection and execution of responses is also imposed [14, 15]. This finding is verified as well in a direct comparison of controlling vs autopilot monitoring in the simulator [16]. In the framework of the present analysis, if monitoring for and responding to failures is regarded as a "task" separate from tacking, then since the operator's attentional resources are limited, the greater workload demands imposed in the control mode than in the monitoring mode would predict poorer performance on the added "task"

of failure detection in the former condition.

Whereas, workload differences make a clear prediction of detection differences in the single-task environment, this prediction is not as apparent when the performance of additional tasks is required. A common result emanating from much dual-task research is that tasks that are in themselves simpler or less loading are at the same time more vulnerable to performance decrements in a dual-task environment, as more demanding paired tasks "capture" a greater proportion of available attentional resources [17]. In the current context, monitoring - the simpler task - should be more vulnerable to additional dual-task requirements than controlling.

Furthermore, to the extent that failure detection while tracking is dependent upon the processing of information integral to the tracking task, then the quality of this information -- and, therefore, the quality of detection itself -- will be preserved as tracking performance is guarded in the face of competing secondary task demands. In contrast, the quality of visual information available in monitoring will be predicted by this view to deteriorate; rendering detection while monitoring more vulnerable to loading than while controlling.

Summary

The implications of the preceding theoretical analysis are complex. In summarizing, two attributes of the controlling mode may be identified that would seemingly facilitate failure detection: a greater consistency of the internal model of the system, and a greater number of channels available upon which to base failure detection decisions. At the same time, the latter advantage may be mitigated to the extent that: (a) adaptation takes place reducing the strength of a visual error signal and, (b) proprioceptive sensitivity is less than visual. In comparison the monitoring mode is also characterized by two attributes that could facilitate detections: a greater "strength" of the visual

signal (if adaptation by an autopilot does not take place) and a lower level of workload.

Finally, it is argued that any advantage of monitoring over controlling attributable to workload differences might itself be dissipated as the competition for attentional resources is increased by imposing concurrent tasks. Clearly this interplay of factors is sufficiently complex to prohibit precise predictions concerning the superiority of one mode over the other. It does, however, facilitate a clearer identification of the nature of the failure detection task and allows predictions to be formulated concerning the differential effect of variables such as workload or control adaptation on detection performance.

In the following experiment, independent variables of participatory mode and task workload were manipulated to determine their effect on detection. Analysis techniques were then employed in an effort to identify further the nature of the processes operating in detection performance.

METHOD

Subjects

The subjects were five right-handed male university students enrolled in basic flight training courses at the Institute of Aviation. Subjects were paid at a rate of $2.50 per hour.

Apparatus

The basic experimental equipment included a 3 x 4 inch Hewlett Packard Model 1300 CRT display, a spring-centered, dual-axis tracking hand control (with an index-finger trigger) operated with the other hand, and a Raytheon 704 16-bit digital computer with 24k memory and A/D, D/A interfacing that was used both to generate inputs to the tracking display and to process responses of the subjects. The subject was seated on a chair with two arm rests, one for the tracking hand controller and one for the side-task finger controller. The

subject's eyes were approximately 112 centimeters from the CRT display so that the display substended a visual angle of 3.4°.

Tracking tasks. The primary pursuit-tracking task required the subject to match the position of a cursor with that of a target which followed a semi-predictable two-dimensional path across the display. The target's path was determined by the summation of two non-harmonically relate sinusoids along each axis. The frequencies were: X-axis, .08 and .05; Y-axis, .08 and .05. The position of the following cursor was controlled jointly by the subject's control response and by a band-limited forcing function with a cutoff frequency of .32 Hz for both axes. Thus the two inputs to the system were well differentiated in terms of predictability, bandwidth, and locus of effect (target vs cursor). The control dynamics of the tracking task were of the form $Y_c = \frac{1-\alpha}{s} + \frac{\alpha}{s^2}$ for each axis, where α was the variable parameter used to introduce changes in the system dynamics. These changes, or simulated failures, were introduced by step changes in the acceleration constant α from a normal value of .3, a mixed velocity and acceleration system with a high weighting on the velocity component, to $\alpha = .9$, a system that approximates pure second order dynamics.

As the loading task, the Critical Task [18], was employed. This was displayed horizontally at the bottom of the screen and required the subject to apply force to the spring-loaded finger control in a left-right direction to keep the unstable error cursor centered on the display. The value of the instability constant λ in the dynamics $Y_c = \frac{k\lambda}{s-\lambda}$ was set at a constant subcritical value. Two values ($\lambda = .05$ and $\lambda = 1.0$) were employed on different dual task trials.

Experimental Task

Subjects participated in five experimental sessions of which the first two were devoted entirely to practice on the tracking and detection tasks, and the last three used to generate the experimental data. During the first practice day the subject performed only the two-dimensional pursuit tracking task. In

the manual (MA) condition the subject performed the tracking manually while in the autopilot (AU) condition, his role in the control loop was replaced by a simulated autopilot control dynamics consisting of a pure gain and effective time-delay. The open loop gain was set at a constant value for all subjects, and the time delay value was adjusted for each subject to obtain an error measure in the AU condition equivalent to the operator's performance in the MA condition. This value of time delay was maintained throughout the rest of the experiment. Each trial, MA or AU, lasted 150 seconds.

To give the subjects some experience with the failed condition (i.e., the higher acceleration in the control dynamics), the subject received two trials (one AU and one MA) in which he tracked (or viewed the autopilot tracking) only the failed dynamics. Two demonstration trials were then presented in which the subject tracked in the regular condition, but the onset of each failure was cued by the presentation of a "F" on the screen. The subject was instructed to press the trigger to return the system to normal only upon the detection of the nature of the change. This training period was then followed by 8 regular detection trials (4 AU, 4 MA in alternating order). Each trial contained either 4 or 6 failures so that a total of 20 failures were presented in each mode.

The presentation of the failure was generated by an algorithm that assured random intervals between presentations and allowed the subject sufficient time to establish baseline tracking performance before the onset of the next change. Task logic also insured that changes would only be introduced when system error was below a criterion value. In the absence of this latter precaution, changes would sometimes introduce obvious "jumps" in cursor position.

During these detection trials, the detection decision was recorded by pressing the trigger on the control stick. This response presented a "T" on the screen and returned the system to normal operating conditions via a four-second ramp to the prefailure dynamics. If the subject failed to detect the

change, the system returned to normal after six seconds. This was an interval within which it was assumed, on the basis of pretest data, that responses would correspond to detected failures and not to false alarms. The subjects were told to detect as many changes as possible as quickly as possible.

On the second day (dual-task training) the subject performed the primary tracking task together with a side task, the Critical Task. After a refresher trial in the MA mode, the subject received a series of training trials to practice the side task, first in the AU and then in the MA mode. When acceptable criteria were achieved in the Critical Task and MA tracking individually, the subject then carried out these tasks together with the failure demonstrations, as described above.

Eight more experimental trials were then presented in which the subject performed all three tasks (tracking or monitoring, Critical Task, and failure detection). Two trials were presented in each mode at each level of Critical Task difficulty (λ = 0.5, 1.0). The subject was instructed to "do the side-stick task as efficiently and accurately as possible." The instructions, therefore, clearly defined the side task as the loading task while allowing performance on the tracking and detection tasks to fluctuate in response to covert changes in available attentional resources. These instructions were emphasized by providing subjects with trial by trial feedback on critical task performance. In this manner, workload demands were experimentally <u>manipulated</u>, rather than being passively assessed.

Following the two training days, the final three days, used to generate the data for experimental analysis followed the format of Table 1. The order of presentation of the 12 experimental trials was counter-balanced across subjects and across days within a subject. The task logic, instructions, and experimental procedure was otherwise identical to that on days 1 and 2.

Insert Table 1 about here

ANALYSIS

Assumptions from signal detection theory [19] were employed to account for detection performance in terms not only of the proportion of failures detected (hit rate), but also the number of detection responses made in the absence of failures (false alarms). The signal detection-based sensitivity index reflects changes in both of these values. Some modification of classical signal detection analysis procedures was required because of the undefined nature of the response interval [20]. According to this procedure it is necessary initially to specify the interval following each failure signal to be designated as a "hit" interval.

The data from a number of pretests, in which dynamics did not return to the pre-failure level, indicated that the distribution of subject responses, following signal occurrence, showed a peak at around three seconds and reached a relatively stable baseline by six seconds following a failure. Therefore, six-second intervals were defined as hit intervals, and the measure P(HIT) was simply the number of detection responses falling within the interval divided by the total number of intervals. The remaining duration of the trial was similarly subdivided into six-second false alarm intervals. The measure P(FA) was computed as the number of false alarms divided by the number of false-alarm intervals.

Because of the relatively small number of signals presented, and the questionable applicability of the formal signal detection theory assumptions to the current data, the nonparametric measure of the area under the ROC curve, P(A), was employed as the bias-free measure of sensitivity [19]. Values to this measure were computed from the P(HIT) and P(FA) data by reference to tables in McNicol [21]. This measure produces a score varying from 0 to 1.0 for which 0.5 represents chance performance and 1.0 represents perfect accuracy. Both

Table 1

Within Subject Experimental Design

(Days 3, 4, 5)

		Participatory Mode	
		Auto (AU)	Manual (MA)
Single Task	Tracking	30 failures	30 failures
Dual Task	Tracking and Easy Critical Task	30 failures	30 failures
	Tracking and Difficult Critical Task	30 failures	30 failures

the P(A) measure and the mean and standard deviation of detection latencies were computed at the end of each trial.

Tracking performance. Tracking measures of vector error, vector control position, and Critical Task error were sampled every 60 msec and stored on digital tape for later data analysis. Error and control position were also differentiated to obtain their respective velocity values. In addition, on a fourth channel, the occurrences of failures and responses were recorded. At the end of each trial, the RMS vector error on the primary task and RMS error on the Critical Task (if performed) were computed.

RESULTS

Detection Performance

Table 2 presents the average hit and false alarm probabilities for each subject in each of the 6 conditions. Following the procedure outlined above, these were converted to P(A) measures, and this bias-free accuracy variable, plotted as a joint function of detection latency and condition, is shown in Figures 2 and 3. Because considerable experimental literature indicates that adding a secondary task, and increasing its difficulty may have qualitatively different effects on primary task performance, [22, 23]. Figure 2 contrasts single with dual-task detection (the mean values of the two critical task conditions), while Figure 3 shows the effect on detection of the Critical Task difficulty level (e.g., $\lambda = .5$ vs. $\lambda = 1$).

Insert Table 2 and Figures 2 and 3 about here

The rationale for the joint speed-accuracy representation of Figures 2 and 3 is that these two variables represent different manifestations of an underlying

Table 2

Detection Data

Subject		MA Single task	MA λ= .5	MA λ= 1.0	AU Single task	AU λ = .5	AU λ = 1.0
C.W.	P(H)	.406	.267	.267	.733	.700	.467
	P(FA)	.071	.060	.080	.130	.100	.200
E.L.	P(H)	.767	.533	.567	.833	.533	.533
	P(FA)	.050	.150	.120	.120	.190	.150
R.S.	P(H)	.467	.433	.567	.533	.458	.423
	P(FA)	.270	.200	.120	.280	.166	.255
T.O.	P(H)	.267	.375	.200	.533	.700	.429
	P(FA)	.110	.133	.170	.120	.110	.186
M A.	P(H)	.500	.400	.167	.833	.800	.750
	P(FA)	.075	.075	.073	.058	.090	.090

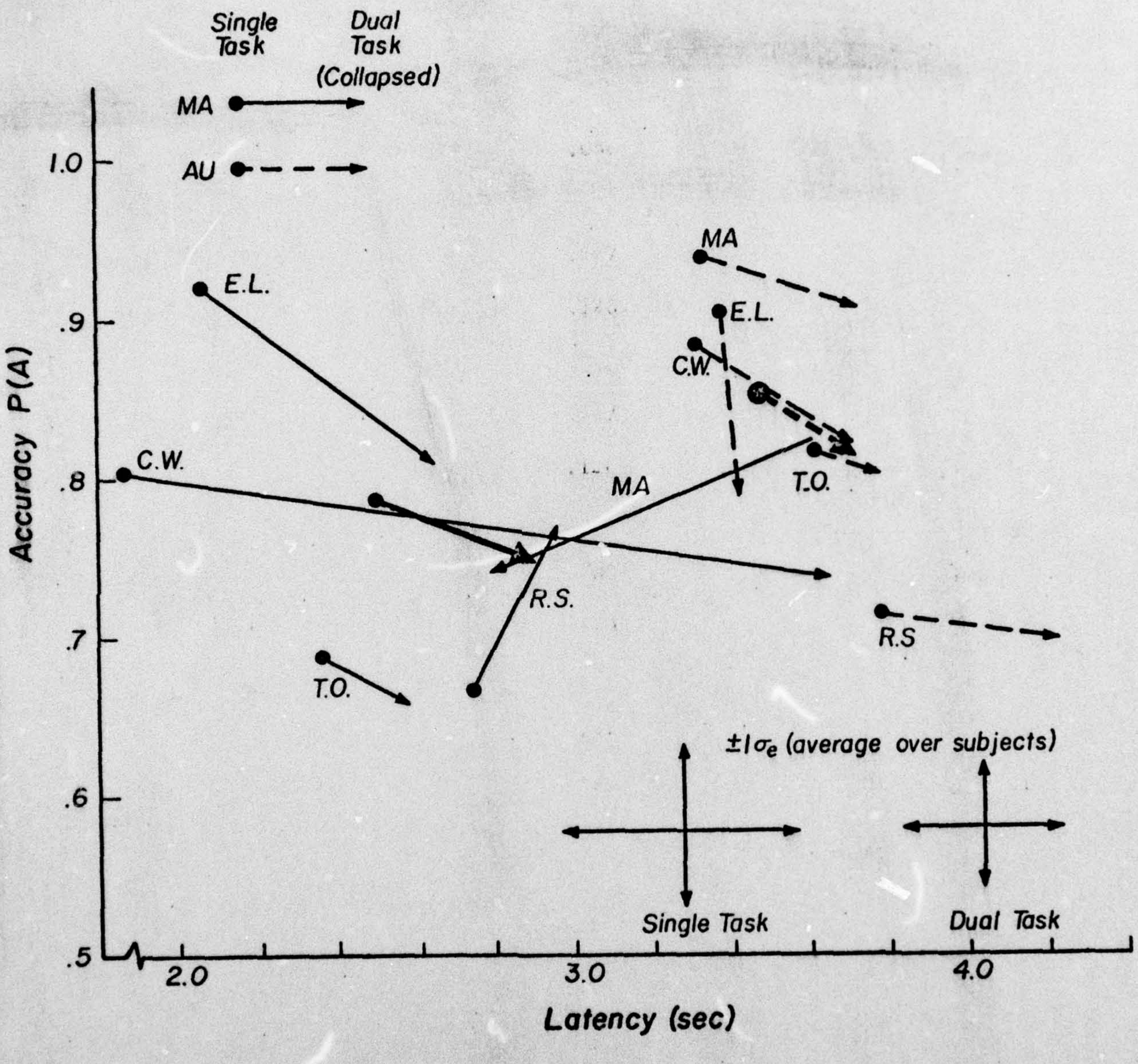

Figure 2: Effect of participatory mode and secondary task performance on detection accuracy and latency. Mean subject trend (heavy lines). Individual subjects (thin lines).

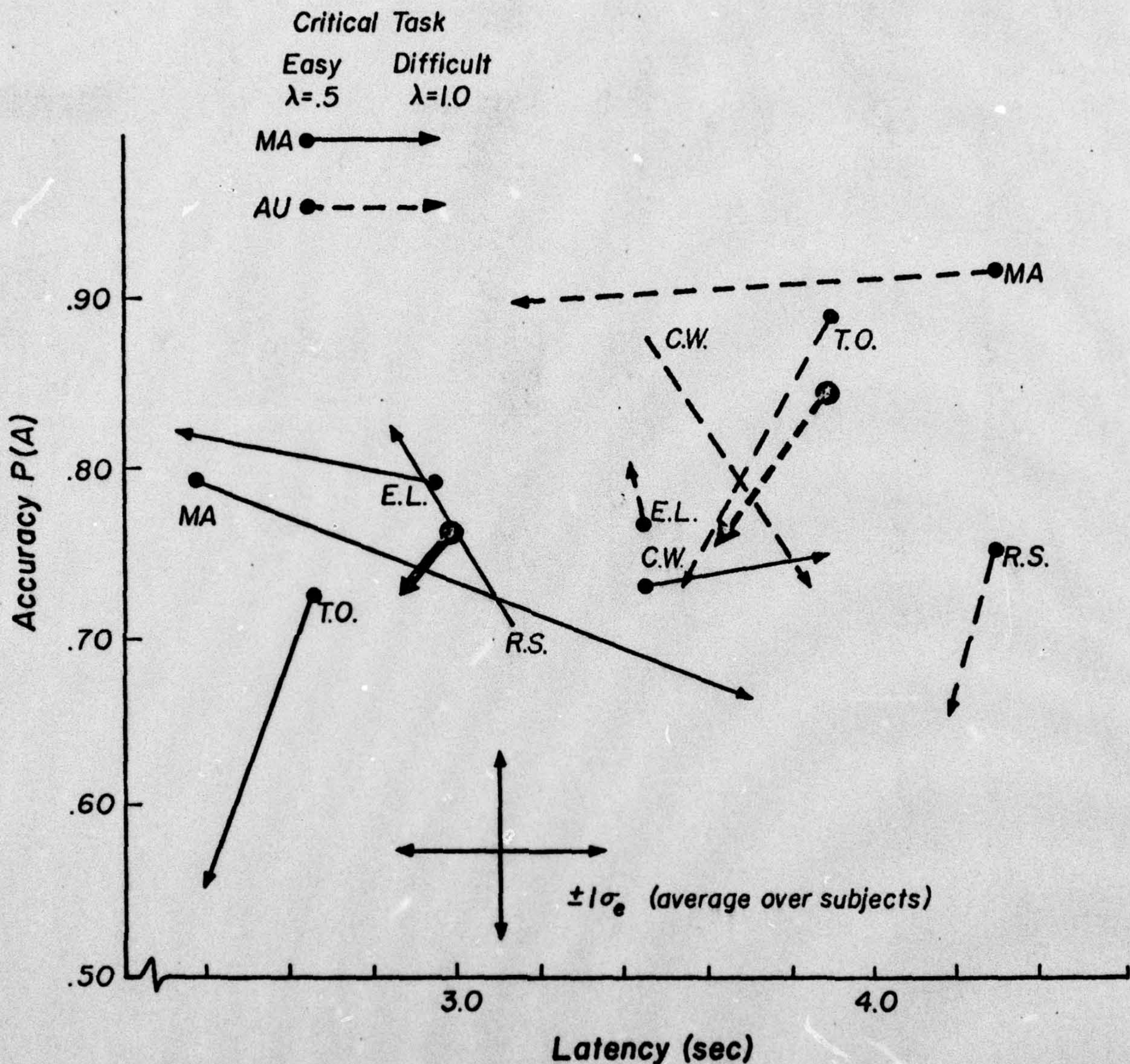

Figure 3: Effects of participatory mode and secondary task difficulty on detection accuracy and latency.

performance metric. In any effort to compare "performance" across conditions, the joint implications of speed and accuracy must be taken into account [24, 25]. For example, a condition that produces a high accuracy of responding might do so at such a prolonged latency that the utility of that decision in a real-world context is less than that of a more rapid decision with slightly lower expected accuracy. Furthermore the underlying model as proposed in the introduction that information is integrated over time until a decision criterion is reached suggests that speed and accuracy may be "traded off" by manipulating the decision criterion. This tradeoff presents another justification for this joint representation.

In Figures 2 and 3, "good" performance (fast and accurate) is represented in the upper left hand regions, while poor performance is in the lower right. In an orthogonal direction, shifts in bias for speed vs. accuracy correspond to movement between the lower left (speed bias) and upper right (accuracy bias) in the space. These shifts may be inferred to relate to variations in setting of the decision making criterion. The lighter vectors portray the data of the five subjects, while the heavier ones represent the mean trend, with each end of the vector corresponding to the bivariate mean of the five individual subject data points. Because of the importance of viewing individual subject data, the brackets below represent the average magnitude, across subjects of ± 1 standard error confidence estimates along both the latency and accuracy axes.[1] By this representation, it is possible to view simultaneously the trend of subject's behavior in the speed-accuracy space, the extent to which the trend typifies the behavior of all subjects and, through the confidence brackets the reliability of the trends shown by individual subjects.

Undoubtedly the most noteworthy effect in Figure 2 is the increase in response latency from the MA to the AU condition ($t = 5.84$, $p < .001$). While this increase in latency for the AU mode is consistent for all subjects, it is

accompanied by an increase in response accurately that is less pronounced and is only evident in both single and dual task conditions for three of the five subjects (C.W., T.O., and M.A.) Thus the prevailing trend induced by shifting from the MA to AU mode appears to be a shift in bias to the upper right in the speed accuracy space: towards slightly more accurate, but considerably slower detection.

The chance in performance induced by the requirement to perform the additional Critical Task (Figure 2) is in the direction of poorer performance: increased latency and decreased accuracy. This is a trend predictable from the assumption of limited operator processing resources. The mean vectors indicate that this trend is of about the same magnitude, and in the same direction for both the AU and MA modes, although it is considerably less consistent across subjects in the MA condition, with two subjects (R.S. and M.A.) showing vectors that do not run in the predicted direction. R.S. shows an accuracy increase, and M.A. a latency decrease with Critical Task performance.

In Figure 3, presenting the effect of increasing Critical Task difficulty on detection performance, the mean trend in both modes shows a decrease in accuracy. However, this is only consistent across subjects in the AU mode (with minor exception of E. L. who shows a minimal accuracy increase). Furthermore, in both modes the increase in Critical Task difficulty appears to lead to a slight decrease in average response latency. Subjects M.A. (MA mode only) and C.W. are exceptions here.

While the trends in neither data set of Figure 3 are striking, the finding of interest is perhaps the very fact that trends are not pronounced when they might otherwise have been expected with increased loading task difficulty. The reason for this expectation is evident from Figure 4 which presents the effects of the workload manipulation on tracking performance. It can be seen here in the

MA mode that tracking performance deteriorates equally as the critical task is added ($t = 3.05$, $p < .01$) <u>and</u> as its difficulty increases ($t = 1.71$, $p < .05$) (the manipulations corresponding to those portrayed in Figures 2 and 3 respectively). Even though there is a slight increase in Critical Task error with λ, it is safe to conclude that increasing Critical Task difficulty leaves fewer processing resources available to devote to performance of the tracking task. Yet in the MA mode, this diminuition of resources does not appear to reduce detection performance in any marked fashion for other than subject M.A. Similarly in the AU mode, the trend of detection performance with increasing λ can <u>not</u> be described as an unambiguous decrease in performance, but rather as a shift in bias to fast inaccurate responding (Figure 3).

Insert Figure 4 about here

<u>Information Utilized in Detection</u>

In the introduction, two hypotheses were proposed to predict why detection might be superior in the MA over the AU mode, and two were proposed predicting AU superiority. The results presented in Figures 2 and 3 and suggest that manual detection was generally superior, and the data were analyzed in detail to determine their consistency with the second hypothesis proposed for manual superiority i.e., the role of the added proprioceptive channel. In this endeavor, two analysis techniques were pursued to identify the cues employed and provide insight into the nature of the detection process. (1) Ensemble averages of display and control variables were constructed primarily to determine the existence of failure "signals": time varying characteristics of the variables, time-locked to failures and increasing during the post-failure interval (Figure 5).Separate averages were constructed for hit and miss trials in each condition. (2) Multiple regression techniques were employed to determine what characteristics of the signal and response were the best predictors of detection

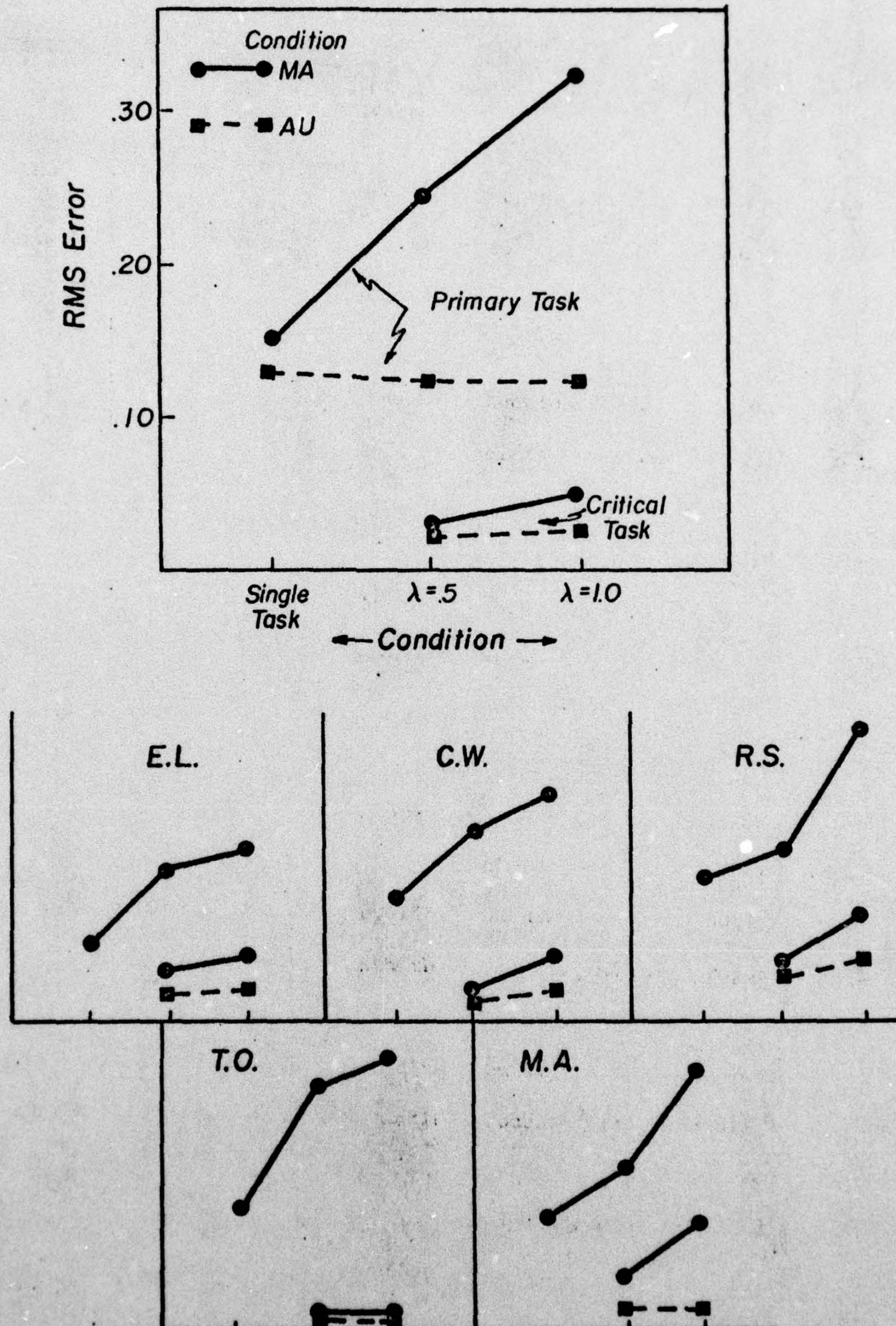

Figure 4: Effects of Critical Task difficulty on performance of primary tracking and of Critical Task. Subject means (top). Individual subjects (bottom).

latency. As predictor variables error, error velocity, control velocity and cursor velocity were sampled at the instant of failure, at .6, 1.2, 1.8, 2.4, 3.6, and 4.8 seconds after the failure. The results of this analysis are presented in Table 3. The results of these two techniques will be discussed as they bear upon the question of the cues utilized for detection.

Insert Figure 5 and Table 3 about here

The assumption that AU detection is based upon the error information is born out by the single task ensemble of averages presented in Figure 5. Clearly a transient error in displayed increase is produced by the failure (information is available to the decision center) and the difference between detected failures (solid lines) and undetected ones (dotted lines) is consistent with the view that undetected failures resulted from a smaller visual, error-based "signal." This analysis is corroborated by the multiple regression data (Table 3). The negative value for the best predictor of AU latency suggests that larger error signals at .6 second latency are associated with faster responses. Similarly the second predictor variable, error velocity, is also associated with latency in such a way as to suggest that increases in its magnitude serve as a signal to shorten detection latency.

Turning to the MA condition in Figure 5 the role played by display error is again evident. Displayed error increases following the failure, and therefore is available as a signal (although it apparently does not increase differentially between hit and miss trials). Furthermore from Table 3 the best predictor of response latency is again the error variable, this time sampled at 1.2 seconds post-failure.

The fact that the increase in the average hit and miss error traces is greater in the AU than the MA condition suggests that in the MA mode the operator

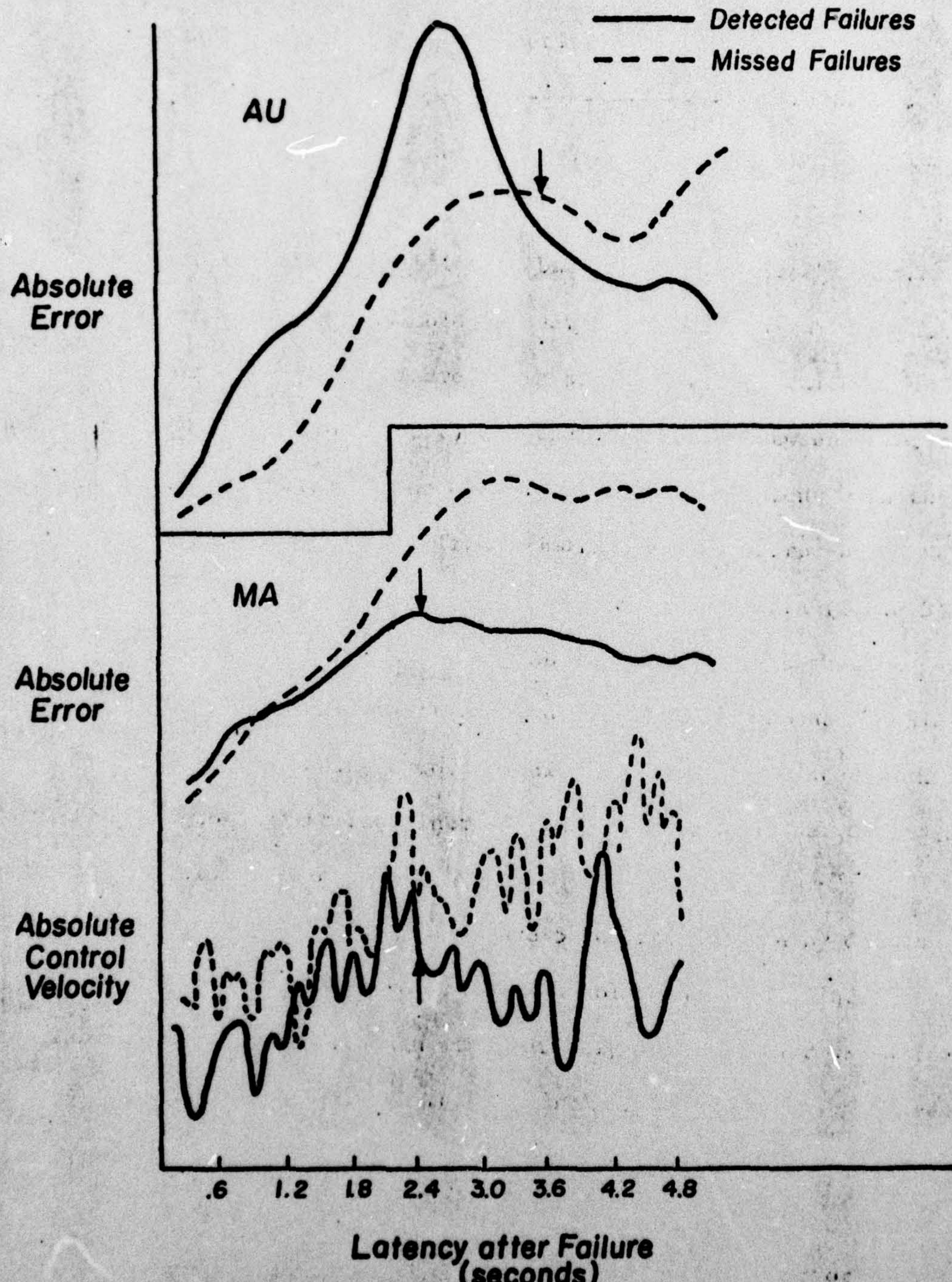

Figure 5: Ensemble averages of error and control velocity for single task AU (top) and MA (bottom) conditions. Vertical arrow indicates mean detection latency.

Table 3

Multiple Regression on Response Latency[a]

Order of Prediction Variable	MA Variable Name	MA Multiple r	MA Partial r	AU Variable Name	AU Multiple r	AU Partial r
1	Error 1.2 sec.	.310	-.310	Error .6 sec	.357	-.357
2	Control vel. 0.6	.470	-.170	Error Velocity .6 sec	.490	-.309
3	Error Velocity 1.2 sec	.521	-.251	Error Velocity 2.4 sec	.524	-.213

[a]Predictor variables were excluded from the table if they occurred at latencies equal to or greater than the mean detection latencies

is performing some sort of control adaptation to the new post-failure plant dynamics--an adaptation directed to bring error to its prefailure level. According to predictions of the crossover model [26], the increase in system order produced by the failure requires the operator to develop greater lead, differentiate the error value and produce a response velocity of higher average value. Thus to the extent that adaptation is carried out, control velocity should increase. Furthermore, even if adaptation is not the linear response predicted by the crossover model, but represents instead a time-optimal bang-bang response [27], the later strategy should still produce an increased control velocity.

In the data presented in Figure 5 a distinct increase in control velocity is visible following failure, supporting the view that adaptation was carried out. This observation supports the hypothesis that MA superiority is based upon the added proprioceptive channel of information, since the figure indicates clearly that this information (manifest in the increasing control velocity) is available prior to detection, as a channel to the decision maker of Figure 1. A further corroberation of the use of this channel is found in the multiple regression analysis that identifies control velocity as the second predictor of response latency. The level of correlation, although modest, is found to be consistently negative with latency, at the .6 and 1.2 second post-failure time points.

<u>Latency distribution</u>. The ensemble average and multiple regression analysis of the single task data suggest that more rapid MA detection may be attributed to the adaptation-related proprioceptive information channel that becomes available to the decision center within 1-2 seconds following the failure occurrence. This interpretation receives more direct support from an analysis of the distribution of response latencies. In the current data, these distributions for all MA conditions were highly skewed in a positive direction, while those of the AU conditions were approximately symmetrical. The latency distributions were

transformed to cumulative probability distributions portraying the relative number or probability of failures detected, as a function of latency after failure (Figure 6). Lappin [28] has argued that a similar representation of his reaction time data -- the latency operating characteristics(sometimes referred to as the cumulative accuracy function) -- may provide evidence bearing upon the time-dependent processes involved in detection: the integration of evidence over time. Following Lappin's approach, it is argued here that the data of Figure 6 may be interpreted as follows: Since detection latencies are assumed to reflect, in part the instant at which the sampled evidence excedes the decision criterion, the extent to which accuracy increases as the criterion moves to longer latencies (through between and within subject variation) is a reflection of the rate of accumulation by the decision center of failure-related information in the post-failure interval. Thus the slope of a function of relative accuracy vs. latency represents the rate at which perceptional evidence becomes available, while the level of the function or intercept represents the overall quality of that information.

Insert Figure 6 about here

According to this interpretation, three important characteristics are evident concerning the single and dual ($\lambda = 1.0$) task data of Figure 6: (1) Both of the MA functions indicate the presence of a distinct discontinuity in the rate of accumulation of evidence, this discontinuity occurring at approximately 1-1.5 seconds post-failure. In a manner consistent with the earlier discussion, it may be argued that the steeper early growth rate reflects the added availability of the proprioceptive signal during the initial 1-2 seconds of adaptation, while the shallower slope following represents the integration of evidence from the visually displayed error signal. (2) The two AU traces, while not strictly

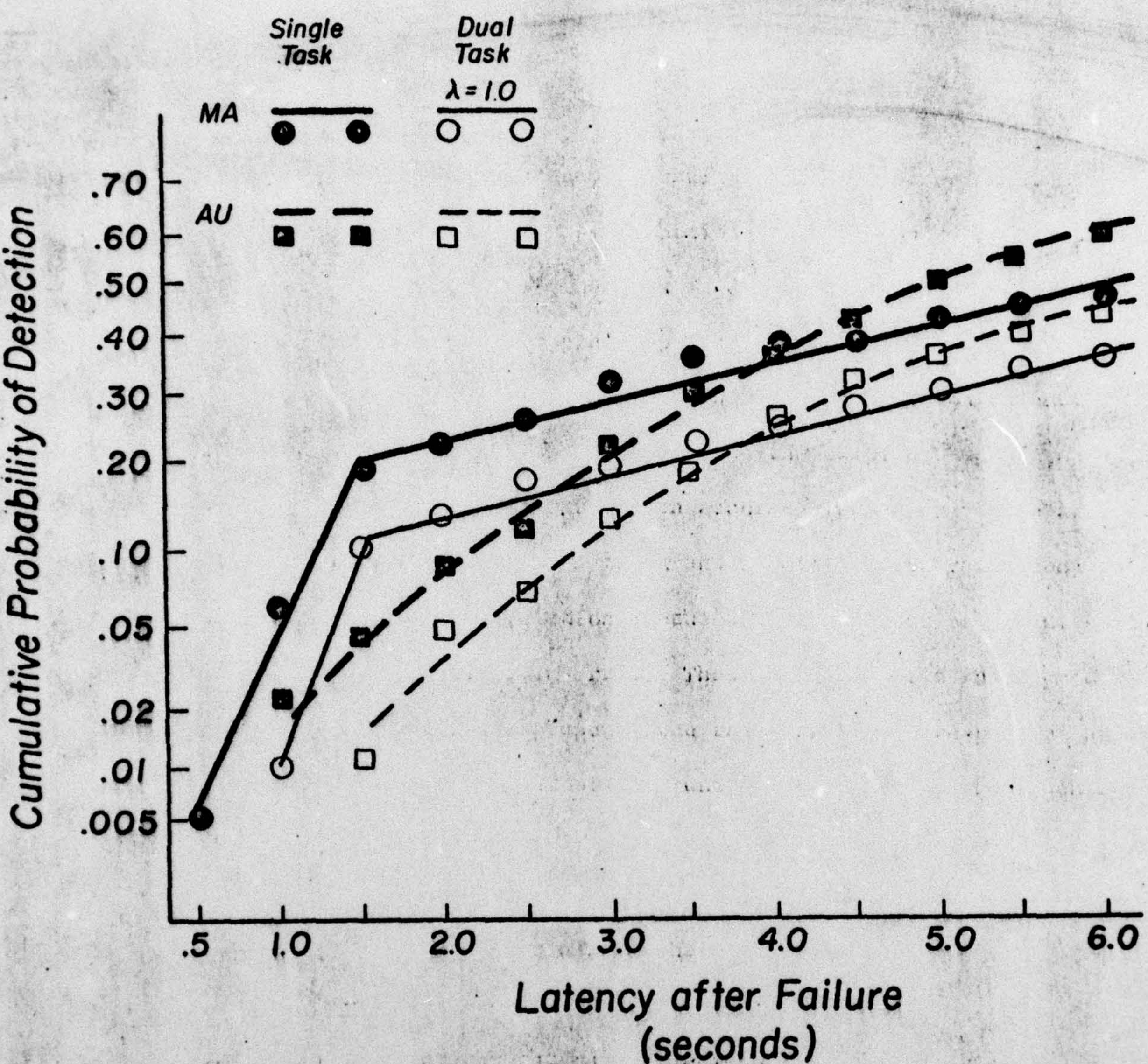

Figure 6: Cumulative probability distribution of detection latencies for single and dual ($\lambda = 1.0$) task AU and MA conditions.

linear, fail to show the abrupt discontinuity of the MA conditions, and thus seemingly represent a uniform underlying process. This process, presumably the integration of displayed visual information accumulates evidence at a <u>faster rate</u> (steeper slope) than is evident in the later, visual portion of the MA mode. This interpretation is consistent with the smaller display error profiles in the MA as compared to the AU conditions (Figure 5), and with the assumption that any adaptation in the MA condition should in fact reduce the magnitude of the visual error signal. With less information therefore available, the integration of this information will procede at a slower rate. (3) In both MA and AU modes, the data of the dual task condition lies below, but closely parallel to the corresponding single task value (intercept shift). This suggests that Critical Task performance, while affecting the overall quality of perceptual data in identical manner for both modes, does not affect the rate of its sampling or acquisition.

<u>Role of Workload</u>. The concept of workload--the task imposed demand for the limited processing resources of the operator--is relevant both as a potential source of AU detection superiority, proposed in the introduction (but not shown by the data) and as it concerns the affect of the added requirement of Critical Task performance on detection. Concerning the first issue it is appropriate to ask why the added workload of controlling the primary tracking task apparently did not hinder the detection of failures in this task, relative to the AU mode when this controlling function was not required and the subject's <u>only</u> task was to monitor the visual display for failures. Two answers may be provided. (1) The role of both error signal and proprioceptive channels demonstrated in MA detection suggests that the very same mental operations that might on the one hand be argued to increase the competition for resources with the detection requirement are also the same ones that are integrally involved in the MA detection process. These operations then may function in <u>cooperation with</u>

the detection process, rather than in competition for the resources upon which detection depends. (2) A different accounting for the absence of a workload deterioration effect can be proposed in terms of the nature of the processing resources themselves. While the resources involved in the failure detection task are primarily related to perceptual and decision making mechanisms, the added resources required by tracking (as opposed to monitoring) concern more directly the response mechanisms. To the extent that perceptual/decision-making processes and response processes draw from different structural pools of processing resources that are not mutually available [22, 23, 29, 30], it is not expected that large interference would be evident between tracking and detection.

This argument, relating to the lack of interference between structurally different processes can also account for the different effects upon detection and tracking performance of a) introduction of the Critical Task (Figure 2), and b) increases in its difficulty (Figure 3). Critical Task introduction adds both a new display element (demand for perceptual resources), as well as new response demands required by the left hand control manipulation (demand for response resources). Thus the expected decrease in both the concurrent tracking performance (Figure 4), and in detection performance (Figure 2) can be predicted, as their respective processing resources are both depleted by the added perceptual and response demands. However, the increase in λ produces primarily an added demand on the availability of resources at the response stages of processing, since the perceptual nature of the Critical Task is little altered (Critical Task error changed little with λ), but a greater motor involvement is required. To the extent that the increase in λ thus depletes response related resources, but not perceptual ones, greater deterioration will be evident in the response loading task (Tracking) than in the perceptual one (Detection).

SUMMARY AND CONCLUSIONS

The major results can be briefly summarized as follows:

1. Detection of step increases in system order when the operator remains in the control loop (MA mode) is considerably faster and only slightly less accurate than when he is removed (AU mode). This finding of MA superiority supports the earlier conclusions of Young (1969).

2. The extent of this superiority did not diminish as the Critical Task was added or as its difficulty was increased by raising the subcritical value of λ: An interaction between participatory mode and workload was not obtained.

3. The effect of adding the Critical Task was to reduce detection performance in both modes, but performance was little altered with increasing λ.

4. Within the framework of the model presented, converging evidence from multiple regression, ensemble averaging and latency distribution data was presented suggesting that the cause of MA superiority was the added proprioceptive information, resulting from control adaptation and available for the first few seconds following the failure. This information, when coupled with the displayed visual information allowed a rapid initial aggregation of evidence in the MA mode, yielding short latency detections. However the availability of the proprioceptive adaptation information was short-lived, due perhaps to the transient memory for the proprioceptive standard. Once gone, the now-adapted visual error signal continued to provide evidence of lesser strength, accumulated at a slower rate than the non-adapting AU condition. Therefore the mean and model latencies of AU detections were longer, but overall detection was slightly more accurate.

5. The role of task workload in affecting detection performance was seemingly only evident as perceptual load was increased (adding the require-

ment of processing the Critical Task display) and not as additional demands were placed upon the response systems, either through tracking the primary task, or through the greater resource demands of the increased loading task difficulty.

6. Concerning the role of a more stable interval model as a cause of MA superiority, it can be argued that this factor probably played a relatively minimal role in influencing the present results. This is because the repeated measures design allowed the same subjects to participate alternately on AU and MA trials. Thus the internal model constructed during MA trials, if superior, was presumably also available on AU trials. On this basis it may be hypothesized that even greater MA superiority might be obtained if participatory mode were manipulated as a between subjects variable.

7. The difference between the findings of Young and of Ephrath regarding MA vs. AU superiority could be accounted for in terms of the current findings. The MA condition in Young's experiment and in the current one were in many respects similar. A step change in system order was imposed following which control adaptation was required entailing a change in general response characteristics (operator describing function) and therefore making available proprioceptive information. Conversely the failure employed by Ephrath, a gradual displacement or bias of the lateral position, is one for which no fundamental change in the operator's transfer function was required to adapt. Therefore proprioceptive channels probably conveyed little if any information relating to the occurrence of a failure.

Finally, some mention should be made concerning the presence of individual differences. To some extent these are inevitable, particularly in a task configuration as complex as the current one, requiring dual task performance in

the AU mode and triple task performance in the MA. Given the subject's flexibility to allocate resources differentially to the two or three tasks, as well as his ability to adapt various criteria on the speed-accuracy detection bias, it is perhaps somewhat surprising that the individual subject data in Figures 2 and 3 are as consistent as they are. Nevertheless, the importance is acknowledged of acquiring more data to replicate and substantiate the trends reported here.

References

1. Sheridan, T. B. Preview of models of the human/monitor/supervisor. In T. B. Sheridan and G. Johannsen, (Eds.), Monitoring Behavior and Supervisory Control, New York: Plenum Press, 1976. 175-180.

2. Young, L. R. On adaptive manual control. IEEE Transactions on Man-Machine Systems, V. MMS-10, 1969, 4, 292-331.

3. Ephrath, A. R. Detection of system failures in multi-axes tasks. Proceedings of the 11th annual NASA-University conference on manual control. NASA TMX, 1975, 62, 151-169.

4. Sheridan, T. B., and Johannsen, G. Monitoring Behavior and Supervisory Control, New York: Plenum Press, 1976.

5. Veldenhysen, W., and Stasen, H. G. The internal model: What does it mean in human control. In T. B. Sheridan and G. Johannsen, (Eds.), Monitoring Behavior and Supervisory Control, New York: Plenum Press, 1976.

6. Pew, R. W. Human perceptual-motor performance. In B. H. Kantowitz, (Ed.), Human information processing tutorials in performance and cognition. New York: Wiley, 1974.

7. Curry, R. E., and Gai, E. G. Detection of random process failures by human monitors. In T. B. Sheridan and G. Johannsen (Eds.), Monitoring Behavior and Supervisory Control, New York: Plenum Press, 1976. 205-220.

8. Miller, D. C., and Elkind, J. J. The adaptive response of the human controller to sudden changes in controlled process dynamics. IEEE Transactions Human Factors in Electronics, HFE, 1967, 8, (3), 218-223.

9. Gould, E. E., and Fu, K. S. Adaptive model of one human operator in a time-varying control task. Proceedings of the second annual NASA-University conference on manual control, NASA SP, 1966, 128, 65-96.

10. Adams, J., Gopher, D., and Lintern, G. Effects of visual and proprioceptive feedback on motor learning. Journal of Motor Behavior, 1977, 9 (1), 11-22.

11. Jordan, T. C. Characteristics of visual and proprioceptive response times in the learning of a motor skill. Quarterly Journal of Experimental Psychology, 1972, 24, 536-543.

12. Klein, R. M., and Posner, M. I. Attention to visual and kinesthetic components of skills. Brain Research, 1974, 71, 401-411.

13. McDonnell, J. D. A preliminary study of human operator behavior following step changes in the controlled element. IEEE Transactions on Human Factors in Electronics, HFE Tech. 1966, 125-129.

14. Keele, S. W. Attention and human performance. California: Goodyear Publishing Company, 1973.

15. Kerr, B. Processing demands during mental operations. Memory and Cognition, 1973, 1, 401-412.

16. Johannsen, G., Pfendler, C., and Stein, W. Human performance and work-load in simulated landing-approaches with autopilot-failures. In T. B. Sheridan and G. Johannsen, (Eds.), Monitoring Behavior and Supervisory Control, New York: Plenum Press, 1976. 83-96

17. Welford, A. T. Fundamentals of Skill. London: Metheun, 1968.

18. Jex, R., McDonnell, J. D., and Phatak, A. V. A "critical" tracking task for manual control research. IEEE Transactions on Human Factors in Electronics, 1966, 7(4).

19. Green, D. M., and Swets, J. A. Signal detection theory and psychophysics, New York: John Wiley and Sons, Inc., 1966.

20. Watson, C. S., and Nichols, T. L. Detectability of auditory signals presented without defined observation intervals. Journal of Acoustic Society of America, 1976, 59, (3).

21. McNicol, D., A primer of signal detection theory, London: George Allen and Irwin Ltd., 1972.

22. Kantowitz, B. H., and Knight, S. L. Testing tapping time-sharing: II. Auditory secondary task. Acta Psychologica, 1967, 40, 343-362.

23. Wickens, C. D., Israel, J., and Donchin, E. The Event Related Cortical Potential as an Index of Task Workload. Proceedings of the Annual Meeting of the Human Factors Society, San Francisco, Calif.: 1977.

24. Pachella, R. The interpretation of reaction time in information processing research. In B. Kantowitz (Ed.), Human Information Processing. Hillsdale, N.J.: Lawrence Erlbaum, 1974.

25. Pew, R. W. The speed accuracy operating characteristic. Attention and Performance II. Acta Psychologica, 1969, 30, 16-26.

26. McRuer, D. T. and Jex, H. R. A review of quasi-linear pilot models. IEEE Transactions on Human Factors in Electronics, 1965, HFE-6, 62-65.

27. Weir, D. H. and Phatac, A. V. Model of human operator response to step transitions in controlled element dynamics. In Second Annual Conference on Manual Control. 1966. NASA SP-128, 65-83.

28. Lappin, J. and Disch, K. The Latency operating characteristic: I. Effects of stimulus probability. Journal of Experimental Psychology, 1972, 92, 419-427.

29. Sanders, A. S. Position paper on human workload. NATO Conference on workload measurement. Mati Greece, September 1977.

30. Wickens, C. D. Position paper on human workload. NATO Conference on workload measurement. Mati Greece, September 1977.

Footnotes

[1]For the accuracy data, standard error of proportion scores were computed for hit and false alarm probabilities by the formula $S_p = \frac{P\ (1 - p)}{N}$. Standard-error confidence brackets were then computed for the accuracy measure P(A) by determining the maximum and minimum accuracy values obtainable when P(Hit) and P(False alarm) were within one standard error of their respective estimated values, e.g., maximum accuracy: $P(H) + 1S_p$, $P(FA) - 1S_p$; minimum accuracy: $P(H) - 1S_p$, $P(FA) + 1S_p$.